Rotation, Rescaling and Occlusion Invariant Object Retrieval

Michela Lecca and Stefano Messelodi
Fondazione Bruno Kessler - IRST
via Sommarive 18, 38050 Povo, Trento, Italy
{lecca, messelod}@itc.it

Abstract

This paper presents a new approach for rotation, rescaling and occlusion invariant retrieval of the objects of a given database D. The objects are represented by means of many 2D views and each of them is occluded by several half-planes. The remaining visible parts (*linear cuts*) as well as the whole views are stored in a new database D' and described by low-level features. Given a portion R of an image, the retrieval of the most similar object is done by generating some linear cuts of R, and by comparing their descriptors with those of the elements of D'. Some heuristic rules regarding visual similarity and geometric properties of the objects in the database drive this process. In the case R is recognized as an object partially occluded, a strategy for the reconstruction of the whole shape of R is also presented.

The tests carried out on synthetic and real-world datasets showed good performances both in recognition and in reconstruction accuracy.

1 Introduction

The diffusion of the digital photo- and video- cameras in the daily life, the rapid increase of the visual materials, their simple and cheap availability, the development of efficient storage devices, make the use of automatic image retrieval systems necessary to manage, index and organize the visual documents in easy and fast way [5].

A content-based image retrieval (CBIR) system is a tool for browsing and searching images from a large database. The system input is a digital image (query) and the output is the database elements or a subset of them ranked by similarity with respect to the query. The similarity is defined as distance between some features describing the database images and the query. Examples of CBIR systems are COMPASS [3], Virage [11], QBIC [7], VisualSEEk [17].

The object retrieval systems are CBIR systems for browsing and searching for objects in a large database of digital images. In this case, the query is a region of a digital image and the output is the database objects or a subset of them ordered by similarity respect to the query. Due to their many applications in the e-commerce and in the automatic indexing of digital images, the development of such systems is an attractive topic in Computer Vision and especially in Image Understanding. An example of commercial application is the search for web catalog products visually similar to a given input. Moreover, these retrieval engines can be inserted in systems for the recognition of objects in digital images

to provide an automatic indexing and a high-level description of images [10]. Since the objects can appear in various accidental circumstances that modify their appearance (like perspective transformations, affinities, occlusions or changes of illuminant), the main bottleneck is to find a description of the objects, able to permit a stable recognition under a large number of conditions.

The most popular approaches describe the objects and the image portion by means of *local features*, like pixel intensity, corners, high curvature points, edge fragments [13], [9], [16], [6]. Other methods describe the objects and the image region by some subsets of pixels [12], or by dividing the object image into many regular parts, for example rectangles, and by describing each of them by global features, [8], [4]. For objects consisting of distinguishable parts arranged in a fixed spatial configuration, models composed by a description of the parts along with information about their spatial relationships or even with a label about their functionality are proposed [14]. Generally, in all these models, the local features are organized in hierarchical structures, for instance trees or graphs, and recognition is reduced to their matching [16], [18].

The choice of the local features for the object description, their extraction from the image, and the matching algorithm for the recognition are often computationally expensive, requiring generally an analysis pixel by pixel or in windows centered in each pixel, and even more for finding the feature correspondences, also when a priori knowledge or heuristic rules are introduced [2].

In this work we propose a method for the retrieval of objects invariant with respect to rotation and/or rescaling and/or occlusions. Differently from the most common approaches, our approach does not use local features for the description of the objects and of the image region to be classified, and the computation of the descriptors as well as the features matching are very fast. The objects are modeled by a set of 2D views and by some occlusions of them by half-planes; given an input region R, the retrieval of the most similar object is done by generating some occlusions of R by half-planes and by comparing them with those of the objects. An algorithm for the reconstruction of the whole shape of occluded objects is also proposed. The main advantages in the use of our method are the simplicity of the object model, the invariance by rescaling, rotating and/or occlusions, and the restricted user interaction. The main disadvantage is due to the fact that the method needs a large amount of memory space to store the database containing the objects and their cuts, but a distributed architecture can easily solve this problem.

2   Overview on the Method

Our method consists of three parts detailed in the next Sections: (1) object model construction, (2) recognition algorithm, (3) reconstruction of recognized occluded objects.
In our approach, the objects to be recognized are represented by many 2D views and they are stored in a database D. Moreover each view is occluded by several half-planes having different slope and masking different percentage of the view. A new database D' containing the remaining visible parts (*linear cuts*) and D is then built.
To establish if a region R of an input image is a view possibly occluded of an object of D, a set of linear cuts of R is generated and then compared them with the items of D'. Each element of D', R, and each linear cut of it are described by means of a vector of low-level features, encoding information about color, shape and texture. The pairs (C_R, C_{O_v}), where
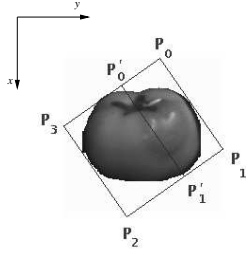
Figure 1: Rectangles for the generation of the linear cuts of a view of an object of the well-known database COIL-100 (`http://www.cs.columbia.edu/CAVE/`).

C_R is a linear cut of R and C_{O_v} is the element of D' having the smallest distance from it (*recognition distance*), are stored in a *candidate list*. R is recognized as an instance of an object O in D if the majority of the linear cuts of R is associated to a linear cut of views of O. Since the descriptors are invariant by rescaling, rotation, and composition of thereof, the system is able to recognize also rotated and/or rescaled partially occluded views. The scale factor of R and its orientation in the image plane with respect to the associated object view are determined and in case of occlusions, they are used for the reconstruction of the whole shape of R. The recognition algorithm is guided by parametric heuristic rules related to geometric properties (area, scale factor) and visual similarity (recognition distance). If R corresponds to an occluded object view, its whole shape is reconstructed and its scale factor and orientation with respect the recognized view are returned. Some thresholds used in the recognition and reconstruction procedures are estimated automatically, while others are user inputs.

Visual similarity between a portion of R and an element of D' is defined as the L^1-distance between their correspondent feature vectors. We call *recognition distance* the minimum distance between R and the items of D'. The image description and the feature matching are performed by the extended version [1] of the content-base image retrieval system COMPASS [3] employed as object classifier. An image portion is described by means of histograms representing intensity, hue, saturation, edge distribution, while the Fourier coefficients and the Li Moments are used to describe its shape. The distributions of hue and saturation are represented also by two dimensional coocurrence matrices. The search process is very efficient, for example retrieving the closest items in a database of 1 million elements takes less than a second on a standard Pentium4 2GHz CPU.

The key point of the recognition algorithm is the computation of the linear cuts of a region Q in $\mathbf{R}^2$ (where Q is an object view in D or the region R). They are generated by occluding Q by $2n$ sheaves of parallel half-planes $\{\Sigma_i\}_{i=1,\dots,m}$ with slopes in $[0, 2\pi]$, where n and m are integer numbers called *generation cut parameters* (briefly GCPs). In particular, for each slope θ_k ($k = 1, \dots, n$), the minimum bounding rectangle of Q with vertices P_0, P_1, P_2, P_3 is built (see Figure 1); a set $\{t_1,\dots,t_m\}$ of real numbers in $(0, 1)$ is fixed and for each i in $\{1,\dots,m\}$ the rectangle with vertices $P_0 + t_i(P_3 - P_0)$, $P_1 + t_i(P_3 - P_1)$, P_2 and P_3 is computed. The Σ_i-linear cuts of Q are then $Q \cap P_0'P_1'P_2P_3$ and $Q - (Q \cap P_0'P_1'P_2P_3)$.

In this work, the slopes are equally spaced in $[0, 2\pi]$, whereas the elements $t_1,\dots,t_m$ are computed by taking in account the size of the database objects, as explained in Section 7.

The GCPs used in the object model construction can be different from those employed in the recognition algorithm and they are set by the user.

3 Object Model Construction

The user inputs are an object database D, the GCPs n and m for the linear cut generation, and two real values p_a and p_b in (0,1), $p_a < p_b$, related to the percentage of the minimum visible area of the linear cuts. We describe each object view O_v of D by low-level features, compute its area $A(O_v)$, occlude O_v by nm half-planes, and then insert O_v and its linear cuts having area in $[p_a A(O_v), p_b A(O_v)]$ in a new database D'. The linear cuts of O_v with area greater than $p_b A(O_v)$ are described and compared with the items of D, and they are put in D' only if the object classifier does not recognize them as instances of O, i.e. the most similar item of D' returned by the object classifier is not a view of O.

The parameters p_a and p_b are used to avoid the generation of too small linear cuts (not significant for recognition) and the insertion in D' of linear cuts very similar to the whole object view. Typically p_a ranges in (0, 0.6], while p_b is close to 1.

4 Heuristic Parameters

The algorithms for the recognition and reconstruction proposed in this work request a threshold Γ on the similarity measure and some heuristic geometric parameters (α_1, α_2, A_m, A_M for the recognition and $\langle E^A_{scale} \rangle$, $\langle E_{angle} \rangle$, a, b, y for the reconstruction).

The values α_1 and α_2 are user inputs and they define the range of variability of the scale factor for the database objects. Denoted by A_{min} and A_{max} the minimum and maximum areas of the elements of D', A_m is given by $\alpha_1^2 A_{min}$ and A_M is given by $\alpha_2^2 A_{max}$.

The values of Γ and $\langle E^A_{scale} \rangle$ are estimated as follows: for each view O_v of an object O of D, s rescaled and rotated versions of O_v are computed. The scale factors and the rotation angles are randomly chosen in $[\alpha_1, \alpha_2]$ and $[0, 2\pi]$ respectively. For each transformed version T of (O_v), a linear cut is randomly generated and classified. If it is recognized as a cut K of a view of O, its scale factor α^* with respect to K is computed as square root between its area and the area of K. The error on scale is defined as the difference $|\alpha^* - \alpha_T|$, where α_T is the scale factor of the transformation T. The distributions of the recognition distances and of the errors on scale factor are computed. Γ is fixed as the 99th percentile of the distribution of the recognition distance, while $\langle E^A_{scale} \rangle$ is the mean value of the scale factors of the recognized objects. In this work, we consider 10 random transformations for each O_v (i.e. $s = 10$).

The value of $\langle E_{angle} \rangle$ measures the error on the angle in the reconstruction and it is defined as function of the GCP n (number of slopes) used for the object model construction, given by

$$\langle E_{angle} \rangle = \frac{1}{2} \frac{\pi}{n}. \tag{1}$$

The values a, b and y are related to the accuracy of the reconstruction and they are set by the user with the constraints $a, b, y > 0$ (see Sections 6 and 7).

5 Recognition Algorithm

The user inputs of this step are a region R of a color image, the GCPs (n_R, m_R) for the generation of the linear cuts of R, and the range $[\alpha_1, \alpha_2]$. The first step is the generation of the set $\{C_R\}$ of linear cuts of R. The cut C_R is a solution only if its area is in the range $[A_m, A_M]$, its scale factor α_R with respect to the most similar item K_R of D' belongs to $[\alpha_1, \alpha_2]$, and its recognition distance $d(C_R)$ is smaller than a threshold Γ. The scale factor α_R is computed by the formula

$$\alpha_R = \sqrt{\frac{A(C_R)}{A(K_R)}} \tag{2}$$

where $A(\cdot)$ indicates the area. The pairs (C_R, K_R) satisfying the conditions enumerated above are stored in the candidate list L and the most frequent object (if any) is returned as solution.

The complexity of the recognition algorithm is proportional to the feature extraction complexity (i.e. $O(A(C_R))$) multiplied the number of elements of the database D' and the number of linear cuts of R compared with the items of D'.

6 Reconstruction Algorithm

The reconstruction algorithm is based on a method of alignment between minimum bounding rectangles. Let (C_R, K_R) be a pair of the candidate list L with scale factor α_R. Let O_v be the object view whose K_R is a cut. The first step to reconstruct the whole shape of R is the determination of the function $\eta : K_R \mapsto C_R$ given by

$$\eta(x) = \frac{1}{\alpha}\mathscr{R}_\psi x + B(C_R) - B(\frac{1}{\alpha}\mathscr{R}_\psi K_R) \tag{3}$$

where x belongs to K_R, $\mathscr{R}$ is the matrix of the rotation of the angle ψ, and $B(\cdot)$ denotes the barycenter. The reconstruction of R is then given by $\overline{\eta}(O_v)$ where $\overline{\eta} : O_v \to \mathbf{R^2}$ is the extension of η on O_v.

The value of α is given by α_R. Let us determine the rotation angle ψ. Let a coordinate system be fixed. Let A, B, C, D and A', B', C', D' be the vertices (in clockwise or anti-clockwise order) of the minimum bounding rectangles of C_R and $\alpha_R^{-1}K_R$ respectively (see Figure 2). Two cases are distinguished:

1. $\text{length}(AB) \neq \text{length}(AD)$ and $\text{length}(A'B') \neq \text{length}(A'D')$. Without loss of generality, we can suppose that $\text{length}(AB) < \text{length}(AD)$ and $\text{length}(A'B') < \text{length}(A'D')$. Since we do not consider the case of flipped objects, the possible transformations of the form (3) that align the rectangles $A'B'C'D'$ and $ABCD$ are two:

$$(A'B'', A'D') \mapsto (AB, AD) \ \text{ or } \ (A'B', A'D') \mapsto (DC, BC) \tag{4}$$

   Thus the angle ψ to be determined can assume the values $\theta_{A'B'} - \theta_{AB}$ or $2\pi + \theta_{A'B'} - \theta_{AB}$, where θ_{AB} and $\theta_{A'B'}$ are the directions of the edges AB and $A'B'$ respectively.

2. $\text{length}(AB) = \text{length}(AD)$ and $\text{length}(A'B') = \text{length}(A'D')$, i.e. the minimum bounding rectangles are squares. In this case there are four possible transformations and the angle ψ can assume the values $\theta_{A'B'} - \theta_{AB} + k\pi$ with $k = 0$, 1/2, 1, 3/2.
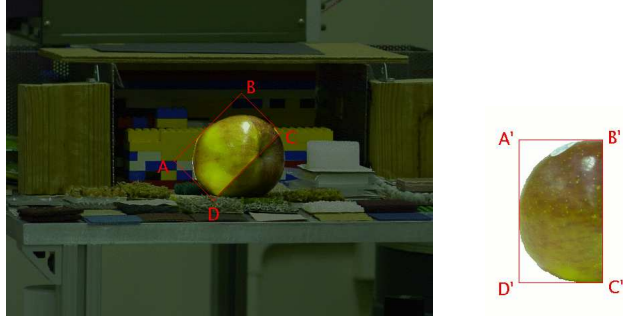
Figure 2: The cut of the apple in the image on left has been put on evidence by making darker the rest of the image. Its bounding rectangle ABCD has to be aligned to the rescaled bounding rectangle A'B'C'D' of the most similar element (on right).

Let $\Psi = \{\psi_i\}_{i \in I}$ be the set of the possible values of ψ.

Since the GCPs used in the recognition algorithm can differ from those employed in the object model construction, and the objects to be recognized can be rotated and/or rescaled, the values of α_R and ψ are affected by error. For this reason, we consider the two sets $S_\alpha = \{\alpha_j\}_{j \in \{0,\dots,a\}}$ of possible scale factors and $S_\psi = \{\psi_{i,k}\}_{i \in I, k \in \{0,\dots,b\}}$ of possible orientations, where

$$\alpha_j = \alpha_R \pm \frac{j}{a} \langle E_{scale} \rangle, \quad \psi_{i,k} = \psi_i \pm \frac{k}{b} \langle E_{angle} \rangle. \tag{5}$$

Then, we compute the functions $\eta_{j,i,k}$ aligning $A'B'C'D'$ and $ABCD$ and thus mapping K_R onto C_R, as in formula (3) with α_j in S_α and ψ_k in S_ψ.

For each pair (C_R, K_R), the functions $\eta_{j,i,k}$ are computed and extended to the functions $\overline{\eta}_{j,i,k} \colon O_v \to \mathbf{R}^2$. All the sets $\overline{\eta}_{j,i,k}(O_v)$ are the possible reconstructions of R. The best reconstruction of R is given by the function $\overline{\eta}^*_{j,i,k}(K_R)$ maximizing the *global overlap index* defined as

$$\frac{A(R \cap \overline{\eta}_{i,j,k}(K_R))}{A(R)} \tag{6}$$

and the *contour overlap index*, defined as the ratio $A(N_1 \cap N_2)/A(N_1 \cup N_2)$ where the regions N_1 and N_2 are defined as follows:

$$N_1 = \partial(\overline{\eta}_{j,i,k}(O_v))^w \cap \partial(\eta(K_R))^w, \quad N_2 = (\partial C_R \cap \partial R)^y. \tag{7}$$

The symbol ∂ denotes the topological boundary of the subsequent region, whereas the superscripts w and y indicate that the correspondent regions in the brackets are grown by w and y pixels respectively.

The parameter w is computed automatically and, as well as y, it is introduced because of the numerical approximations that occur in the computation of the function $\overline{\eta}^*_{i,j,k}$. In fact, the numerical approximations cause an imperfect overlap between the restriction of $\overline{\eta}_{i,j,k}(O_v)$ to $\eta_{i,j,k}(K_R)$ and $\eta_{i,j,k}(K_R)$. The difference between the two borders

$\partial(\overline{\eta}_{j,i,k}(O_v)) \cap \partial(\eta(K_R))$ and $\partial(C_R) \cap \partial(R)$ is generally very small, but it could be significant in the computation of the contour overlap index. For this reason, w is estimated by rounding up the tolerance

$$2\sqrt{\frac{A(\eta_{i,j,k}(K_R) \cap \overline{\eta}_{i,j,k}(O_v))}{A(\eta_{i,j,k}(K_R))}} \tag{8}$$

to the nearest integer.

On the contrary, the parameter y is set by the user and it is related to the accuracy on the reconstruction, that is measured by the overlap index $v = A(R \cap \overline{\eta}^*_{j,i,k}(O_v))/A(R)$. In our tests, we set $y = 2$ pixels.

The complexity of the reconstruction algorithm for a cut C_R is proportional to the product $4|I|abA(C_R)$, where $|I|$ is the cardinality of the set I.

7   Experiments and Conclusions

The section presents the experiments about the recognition and reconstruction performances obtained by running our method on 4 sets of synthetic images and on a set of real-world pictures. The tests on the synthetic data include some studies about the dependency of (*i*) the recognition rate on the GCPs used in model construction and in recognition algorithm, and (*ii*) the reconstruction accuracy on the parameters a and b.

Experiments on the Synthetic Data - Database IKEA-400 consists of 400 images of furniture (400 objects) of IKEA SpA downloaded from the internet catalog
`http://www.ikea.com/ms/it_IT/our_products.html`.

The recognition and reconstruction performances have been tested on 4 datasets S_{15}, S_{25}, S_{35}, S_{45} containing rescaled and/or rotated versions of the objects of IKEA-400 partially occluded by synthetic patches (Figure 3). The subscript in the set names indicates the percentage of occluded area. The scale factor and the rotation angle range in [0.3; 1.5] and [0; 2π] respectively.

For the object model construction, we considered 10 sheaves of half-planes, i.e. 10 slopes equally spaced in [0, 2π]. Since the objects in the database have different size, the number of cuts having the same slope are different for each object and has been fixed by requiring that the minimum distance between two half-plane of the same sheaf is $\Delta = 50$ pixels. Moreover, $p_a = 0.40$ and $p_b = 0.90$. The resulting database IKEA-400' contains 10660 images. In particular, we obtained the following values: $\Gamma = 0.03574$, $\langle E_{scale} \rangle = 0.20044$, $\langle E_{angle} \rangle = 0.15708$, $A_m = 675.75$, and $A_M = 438898.5$ pixels. For the reconstruction we fixed $a = b = 2$.

We submit each visible part of the objects in the 4 test sets to the recognition algorithm with GCPs $n_R = 2$, $m_R = 2$ (case a.) and $n'_R = 5$, $m'_R = 5$ (case b.). Tables 1-a and 1-b show the recognition rate ρ and the reconstruction accuracy, measured by the mean overlap index $\overline{v}$ obtained by setting $a = b = 2$. In case a., the computational time for the recognition is less than 1 second, in case b. it is about 2.6 seconds, but the results are better. The recognition rate ρ is in case b. very satisfactory. For $a = b = 2$, the reconstruction requests about 6 seconds. For case b. we repeated the tests by fixing $a = b = 4$ and we obtained the mean overlap index $\overline{v}'$. There is not a significant difference between the values of $\overline{v}$ and $\overline{v}'$, but for $a = b = 4$, the computational time is longer (about 20 seconds). The times have been measured by running the algorithm on a standard Pentium4 2.80GHz.
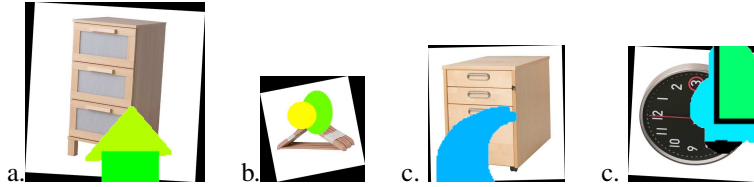
Figure 3: Examples of synthetic test images. Each of them contains a rescaled and/or rotated object view of D partially occluded (occluded area percentages of a. 15%, b. 25%, c. 35 %, and d. 45%).

| **Table a** | | | | **Table b** | | | | | **Table c** | | |
Test Set	ρ	$\overline{v}$		Test Set	ρ	$\overline{v}$	$\overline{v}'$		Test Set	ρ	$\overline{v}$
S_{15}	0.865	0.756		S_{15}	0.983	0.882	0.892		S_{15}	0.980	0.876
S_{25}	0.748	0.725		S_{25}	0.885	0.848	0.852		S_{25}	0.876	0.814
S_{35}	0.798	0.666		S_{35}	0.965	0.815	0.821		S_{35}	0.945	0.799
S_{45}	0.668	0.626		S_{45}	0.955	0.796	0.821		S_{45}	0.918	0.762

Table 1: Recognition rate and mean overlap index for the test sets with objects of IKEA-400. The overlap indices v and v' have been obtained by setting ($a = 2$, $b = 2$) and ($a = 4$, $b = 4$) respectively. Tables a. and b. are referred to the database IKEA-400', whereas the results of Table c. have been obtained by using the database IKEA-400''.

To test the results dependency on the GCPs used for the object model construction, we built an other extension of IKEA-400 by generating cuts with 5 sheaves of half-planes with $\Delta = 100$ pixels. The resulting database (IKEA-400'') contains 5573 images. Table 1-c shows the recognition rate and the mean overlap index obtained by using in the recognition algorithm the GCPs $n_R' = m_R' = 5$ and the parameters $a = b = 2$ for the reconstruction. Also in this case satisfactory results have been obtained.

Experiments on the Real-World Data - The tests on the recognition and reconstruction performances of our method on real-world images have been carried out by considering a database of 8 objects (that we denote with 3DOBJ) along with a ground-truth of 51 pictures containing whole or partially occluded objects of 3DOBJ [15]. This material is freely available on `http://tev.itc.it/DATABASES/objectsPonce.html`. The database 3DOBJ consists of 161 images of 8 objects, each represented by some 2D views. The database 3DOBJ' extended by the cuts has been built by considering 10 sheaves of half-planes, each of them - as well as in the case of the synthetic data - contains a number of half-planes different for each objects. The values of p_a and p_b are 0.20 and 0.90 respectively. The heuristic thresholds have the following values: $\Gamma = 0.02493$, $\langle E_{scale} \rangle = 0.20104$, $\langle E_{angle} \rangle = 0.15708$, $A_m = 442.35$, and $A_M = 250762.5$ pixels. The parameters for the reconstruction are $a = 5$, $b = 7$. We selected by a semi-automatic segmentation the visible parts of the objects in the test images and we classify them by setting $n_R = m_R = 10$. The obtained recognition rate is 80.89%. The most of cases of non recognition are due to the fact that the objects in the test images are differently illuminated with respect to the object views in the database and our features are not invariant by changing the light conditions. The reconstruction is in this case harder, because in the test images, the objects to be recognized appear under different points of view respect to the shots stored in the database. This is the case of the shoe shown in Figure 4. Nevertheless, the mean

Figure 4: The set of views of the shoe in the database D (left) does not contain the pose of the test image on right. In this case, the object is recognized, but a good reconstruction is not possible.
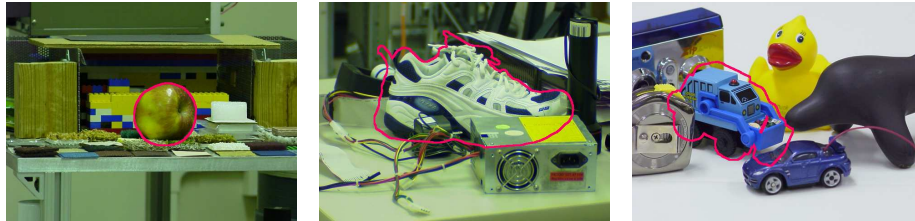

Figure 5: Examples of Reconstruction in Real-World Images.

overlap index is 0.9280. Some results are shown in Figure 5.

Conclusions - The tests illustrated in this Section show that our method performs good both in term of recognition and reconstruction. Therefore our future work includes its use in an object recognition system like [10]. Since the computational times are quite long, our planes comprehend also a parallelization and optimization of the code and the comparison of our approach with other on common databases. Moreover, we will develop a strategy for the automatic estimation of the thresholds that are currently to be fixed by the user, like the GCPs parameters.

References

[1] C. Andreatta. CBIR techniques for object recognition. Technical Report, ITC-irst T04-12-01, December 2004.

[2] S. Berretti, A. Del Bimbo, and E. Vicario. Efficient matching and indexing of graph models in content-based retrieval. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 23(10), 2001.

[3] R. Brunelli and O. Mich. Image retrieval by examples. *IEEE Transactions on Multimedia*, 2(3), 2000.

[4] T. Deselaers, D. Keysers, R. Paredes, E. Vidal, and H. Ney. Local representation for multi-object recognition. In *Deutsche Arbeitsgemeinschaft für Mustererkennung: DAGM 2003*, 2003.

[5] H. Eidenberger. A new perspective on visual information retrieval. In *SPIE Electronic Imaging Symposium, SPIE vol. 5304*, San Jose, 2004.

[6] V. Ferrari, T. Tuytelaars, and L. Van Gool. Simultaneous object recognition and segmentation by image exploration. *International Journal of Computer Vision (IJCV)*, (4), 2006.

[7] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, H. Qian, and other. Query by image and video content: the QBIC system. *Computer*, 28(9), 1995.

[8] G. Fritz, L. Paletta, and H. Bischof. Object representation and recognition from informative local appearances. In *Digital Imaging in Media and Education*, 2004.

[9] F. Krolupper. Recognition of occluded objects using curvature. In *Proc. of The 12th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, 2004.

[10] M. Lecca. Object recognition in color images by the self configuring system MEM-ORI. *International Journal of Signal Processing*, 3(3), 2006.

[11] Philippe Mulhem and Joo Hwee Lim. Symbolic photograph content-based retrieval. In *CIKM '02: Proceedings of the eleventh international conference on Information and knowledge management*, New York, NY, USA, 2002. ACM Press.

[12] S. Obdrzlek and J. Matas. Object recognition using local affine frames on distinguished regions. In *British Machine Vision Conference*, 2002.

[13] M. Reinhold, M. Grzegorzek, and H. Niemann J. Denzler. Appearance-based recognition of 3-d objects by cluttered background and occlusions. *Pattern Recognition*, 38(5), 2005.

[14] E. Rivlin, S. J. Dickinson, and A. Rosenfeld. Recognition by functional parts. *Computer Vision and Image Understanding: CVIU*, 82(2), 1995.

[15] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. 3D object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints. *IJCV*, 66(3), 2006.

[16] A. Shokoufandeh, I. Marsic, and S. J. Dickinson. View-based object recognition using saliency maps. *Image and Computing*, 17, 1999.

[17] J. R. Smith and S.F. Chang. VisualSEEk: A fully automated content-based image query system. In *ACM Multimedia*, 1996.

[18] V. Vilaplana, X. Giro, P. Salembier, and F. Marques. Region-based extraction and analysis of visual object information. In *Proc. of Int. Workshop on Content-Based Multimedia Indexing CBMI 2005*, 2005.